

Utah State University

DigitalCommons@USU

Economic Research Institute Study Papers

Economics and Finance

2000

Measurement Error and the Distribution of Income

L. Dwight Israelsen
Utah State University

James B. McDonald

Follow this and additional works at: <https://digitalcommons.usu.edu/eri>

Recommended Citation

Israelsen, L. Dwight and McDonald, James B., "Measurement Error and the Distribution of Income" (2000).
Economic Research Institute Study Papers. Paper 194.
<https://digitalcommons.usu.edu/eri/194>

This Article is brought to you for free and open access by the Economics and Finance at DigitalCommons@USU. It has been accepted for inclusion in Economic Research Institute Study Papers by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



Economic Research Institute Study Paper

ERI #2000-18

**MEASUREMENT ERROR AND THE DISTRIBUTION
OF INCOME**

by

L. DWIGHT ISRAELSEN

**Department of Economics
Utah State University
3530 Old Main Hill
Logan, UT 84322-3530**

JAMES B. MCDONALD

**Department of Economics
Brigham Young University
Provo, UT 84602**

April 2000

**MEASUREMENT ERROR AND THE DISTRIBUTION
OF INCOME**

L. Dwight Israelsen, Professor

**Department of Economics
Utah State University
3530 Old Main Hill
Logan, UT 84322-3530**

James B. McDonald

**Department of Economics
Brigham Young University
Provo, UT 84602**

The analyses and views reported in this paper are those of the author(s). They are not necessarily endorsed by the Department of Economics or by Utah State University.

Utah State University is committed to the policy that all persons shall have equal access to its programs and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

Information on other titles in this series may be obtained from: Department of Economics, Utah State University, 3530 Old Main Hill, Logan, Utah 84322-3530.

Copyright © 2000 by L. Dwight Israelsen and James B. McDonald. All rights reserved.

Readers may make verbatim copies of this document for noncommercial purposes by any means, provided that this copyright notice appears on all such copies.

MEASUREMENT ERROR AND THE DISTRIBUTION OF INCOME

L. Dwight Israelsen and James B. McDonald

ABSTRACT

A model for measurement error is developed, based on the assumption that measurement error is random, multiplicative, and independent of the level of actual income. Thus, measured income is defined as the product of actual income and measurement error. Flexible parametric forms are utilized to model the distributions of actual income (generalized gamma) and measurement error (inverse generalized gamma). The probability density of measured income is then derived as a generalized beta of the second kind (GB2). Estimation of the parameters of the GB2 (measured income), then allows an estimate to be made of the pdf of actual income, from which corresponding estimated means, variances, Gini coefficients, and Lorenz curves are obtained. An identification problem in the parameters of actual and measured income is solved with additional information as to the average fraction of actual income reported. The implied characteristics of measurement error are also obtained. The procedure is applied to income data from several Latin American economies, and estimates of actual income distribution characteristics are derived from measured income. It is found, in some cases, that when measured income inequality moved in one direction over time, actual income inequality moved in the opposite direction. This finding has important implications for the evaluation of policies designed to affect relative equality in the distribution of income.

MEASUREMENT ERROR AND THE DISTRIBUTION OF INCOME

1. Introduction

During the past twenty years, there has been a renewed interest in the United States and around the world in the measurement of income inequality. In the United States, this development has been due, in part, to the uncertainty surrounding the trend in income inequality in the economy since the 1970s. Part of the uncertainty is a result of the recognition by economists and policymakers that the data from which income distributions are estimated is subject to measurement error. The measurement error has generally been assumed to be associated with underreporting; however, during the past decade concern has also been expressed about the problem of measurement error/evasion due to overreporting of income by low income families wanting to claim the earned income tax credit.

The purpose of the research reported here is to develop a tractable method to estimate the actual distribution of income from the distribution of measured income in the presence of measurement error with certain characteristics. Estimation of the parameters of the distribution of measured income would then allow us to recover an estimate of the probability density function of actual income. From the pdf of actual income, we could then obtain corresponding estimated means, variances, Gini coefficients and Lorenz curves.

This paper expands and improves on our earlier paper (Israelsen, McDonald & Newey, 1984) in which we derived several pdf's for measured income and demonstrated the theoretical possibilities of the methodology. Unfortunately, the distribution functions we derived there were

difficult to estimate empirically, so one goal of the current paper is to derive a pdf of measured income which is both flexible and empirically tractable. We have accomplished that goal and have derived a distribution function for measured income which has been demonstrated to provide an excellent fit to empirical income data. We have applied the methodology to data from Mexico and Latin America, and have found that measurement error of the magnitude that exists in those countries not only can lead to biased estimates of income inequality at a specific time, but can also obscure the actual direction of change in income inequality over time. Indeed, we found that in some cases, between 1980 and 1989, where the measured distribution of income showed an increase in apparent inequality, as indicated by the Gini coefficient, the actual distribution of income based upon our procedure showed a decrease in inequality, and vice versa. Since much of the economic development assistance in these and other countries has the goal of reducing economic inequality, the failure to accurately measure trends in actual inequality can lead to serious mistakes in the formulation of assistance programs and the allocation of resources within such programs.

The implications of this research also extend to the United States and other industrialized countries. One of the puzzles U.S. economists have faced is in explaining the apparent inability of redistribution and incentive programs to reduce income inequality during the past quarter century. It may be that part of that puzzle is obscured by the impact of changes in underreporting on measured income inequality. If the technique outlined here proves to be applicable to U.S. income data, it could have a major impact on economists' assessment of the effectiveness of incomes policies in developed economies.

2. The Model

Several approaches to analyzing the impact of measurement error on observed inequality are possible. The model we have adopted for measured income has been shown to provide a very good fit to observed income distributions in a variety of different applications. The specific model utilized is based upon the assumption that the ratio of reported income to actual income is a random variable which is independent of the level of actual income. Flexible parametric forms are utilized for the distribution of measured and actual income as well as for measurement error. We first define notation and then formally develop the model relating the distributions of actual and measured income.

Let Y and Y^* , respectively, denote measured and actual income. $\lambda = Y/Y^*$ is assumed to be distributed independently of Y^* . Consequently, measured income is related to actual income by $Y = \lambda Y^*$ where λ denotes a multiplicative measurement error. The expected value of λ is the fraction of actual income that is reported. We further assume that

$$Y^* \sim GG(y; a, b, p) \quad (2.1)$$

$$\text{and} \quad \lambda \sim IGG(\lambda; a, \beta, q) \quad (2.2)$$

where GG denotes the generalized gamma distribution defined by the pdf

$$GG(y; a, b, p) = \frac{y^{ap-1} e^{-(y/b)^a}}{b^{ap} \Gamma(p)} \quad (2.3)$$

and IGG denotes the inverse generalized gamma defined by $IGG(\lambda; a, \beta, q) = GG(\lambda; -a, \beta, q)$. In (2.3), a and p are shape parameters with b being a scale parameter.

The selection of a generalized gamma to model actual income includes the gamma, Weibull, and lognormal distributions as special cases. The gamma, Weibull and lognormal do

not permit intersecting Lorenz curves; however, the generalized gamma allows for, but does not impose, intersecting Lorenz curves (Taillie, 1981). The h^{th} order moments of the generalized gamma are given by

$$E(Y^h) = \frac{b^h \Gamma(p+h/a)}{\Gamma(p)} \quad (2.4)$$

and the corresponding Gini coefficient¹ can be expressed as

$$\text{Gini}_{\text{GG}} = \frac{\left[\left(\frac{1}{p} \right) {}_2F_1 \left[1, 2p+1/a; p+1; \frac{1}{2} \right] - \left(\frac{1}{p+1/a} \right) {}_2F_1 \left[1, 2p+1/a; p+1/a; \frac{1}{2} \right] \right]}{[2^{2p+1/a} B(p, p+1/a)]} \quad (2.5)$$

where the ${}_2F_1 []$ denotes a hypergeometric series (Rainville (1960)).

$$\text{The mean of } \lambda \text{ (measurement error) is given by } E(\lambda) = \frac{\beta \Gamma(q-1/a)}{\Gamma(q)}. \quad (2.6)$$

It can be shown that the $\lim_{q \rightarrow \infty} \text{Variance}(\lambda) = 0$.

The probability density of measured income (Y) can be derived from (2.1) and (2.2) as

$$\begin{aligned} f(y) &= \int_0^{\infty} GG(y/\lambda; a, b, p) \text{IGG}(\lambda; a, \beta, q) (d\lambda/\lambda) \\ &= \int_0^{\infty} GG(y; a, \lambda b, p) \text{IGG}(b\lambda; a, b\beta, q) d(\lambda b) \\ &= \int_0^{\infty} GG(y; a, t, p) \text{IGG}(t; a, b\beta, q) d(t) \end{aligned}$$

¹The derivation of the expression for the Gini coefficient can be found in McDonald (1984).

$$= \text{GB2}(y; a, b\beta, p, q) \quad (2.7)$$

where the GB2 denotes the generalized beta of the second kind defined by

$$\text{GB2}(y; a, s, p, q) = \frac{|a| y^{ap-1}}{s^{ap} B(p, q) (1 + (y/s)^a)^{p+q}} \quad (2.8)$$

where s denotes a scale parameter and a , p , and q are shape parameters. The h^{th} order moments of the GB2 in (2.8) can be written as

$$E_{\text{GB2}}(Y^h) = \frac{(s)^h B(p+h/a, q-h/a)}{B(p, q)} \quad (2.9)$$

with the corresponding Gini coefficient being given by

$$\text{Gini}_{\text{GB2}} = \frac{\left[\left(\frac{1}{p} \right) {}_3F_2 \left[\begin{matrix} 1, p+q, 2p+1/a; 1 \\ p+1, 2(p+q); \end{matrix} \right] - \left(\frac{1}{p+1/a} \right) {}_3F_2 \left[\begin{matrix} 1, p+q, 2p+1/a; 1 \\ p+1/a+1, 2(p+q); \end{matrix} \right] \right]}{[2^{2p+1/a} B(p, p+1/a) / B(2q-1/a, 2p+1/a)]} \quad (2.10)$$

McDonald(1984).

The parameters in (2.8) can be estimated using standard maximum likelihood estimation procedures yielding \hat{s} , \hat{a} , \hat{p} , and \hat{q} . Note that the scale parameter in the GB2 () for measured income in (2.7) is equal to the product of the scale parameters b and β from the pdf's for actual income and measurement error. Thus, while we can estimate the product of b and β , we are not able to obtain separate estimates of b and β without additional information. This represents an identification problem. The additional information can take a variety of forms. One such case is in applications where an estimate of the mean level of λ is known.

We now summarize the estimation procedure for the case in which the mean of

measurement error, $\bar{\lambda}$, is known.

- (1) Estimate the parameters of the GB2 ($y; a, s = b\beta, p, q$) in (2.8) using MLE methods to yield \hat{a} , $\hat{s} = (\hat{b}\hat{\beta})$, \hat{p} , and \hat{q} .
- (2) Solve equation (2.6), $\bar{\lambda} = \beta\Gamma(\hat{q}-1/\hat{a})/\Gamma(\hat{q})$ for $\hat{\beta}$ i.e.,

$$\hat{\beta} = \bar{\lambda}\Gamma(\hat{q})/\Gamma(\hat{q}-1/\hat{a}) .$$

- (3) Given an estimate of β , $\hat{\beta}$, obtain an estimate of b ,

$$\hat{b} = \frac{\hat{s}}{\hat{\beta}} = \frac{\hat{b}\hat{\beta}}{\hat{\beta}} .$$

We then have an estimate of the pdf of actual income

$$GG(y^*; \hat{a}, \hat{b}, \hat{p})$$

from which estimated means, variances, Gini coefficients, and Lorenz curves can be obtained.

The implied characteristics of “measurement error” can be obtained by studying the pdf of λ

$$IGG(\lambda; \hat{a}, \hat{\beta}, \hat{q}) .$$

In summary, we note that if actual income is distributed as a generalized gamma and the measurement error is distributed as a particular member of the inverse generalized gamma family, then measured income will be distributed as a generalized beta of the second kind (GB2).

3. Empirical Results

The methodology developed in Section 2 was applied to World Bank data for Argentina, Columbia, Costa Rica, Guatemala, Mexico, Uruguay, and Venezuela for the period 1980-1989, (Psacharopoulos, et al., 1993). In addition to measured income, the World Bank researchers were also able to provide estimates of the mean fraction of income reported. These estimates were found by comparing budget survey data with national income data, and allowed us to solve the identification problem mentioned in Section 2. Table 1 summarizes the results of the income analysis. Note that the mean fraction of actual income reported ($\bar{\lambda}$) ranges from .2691 for Costa Rica, 1989, to .9035 for Mexico, 1989. Note also, that the mean fraction of income reported changes significantly between years for most of the countries in the study. In Argentina, Costa Rica, and Venezuela, the mean fraction of income reported fell over the period, while in Guatemala and Mexico, the mean fraction of income reported increased between years. The mean fraction reported changed only slightly in Columbia and Uruguay. It has been shown that measurement error of this form exaggerates income inequality (Arnold (1980) and Israelsen, McDonald, and Newey (1984)). We find that significant changes in the mean fraction of income reported could cause reported income inequality to move in one direction, while actual income inequality moved in the opposite direction. As can be seen from Table 2, this occurred in several of our included countries. In Columbia and Venezuela, measured income inequality increased over the period, while actual income inequality decreased. In Guatemala and Mexico, on the other hand, measured income inequality decreased, while actual income inequality increased. It is also interesting to note the magnitude of the bias in the change in inequality even in countries where measured and actual inequality moved in the same direction. In Uruguay, the

measured reduction in inequality was twice as great as the actual reduction in inequality, whereas in Argentina, the measured increase in inequality was 7 times as great as the actual increase in inequality.

Figure 1 illustrates the estimated distributions of actual and measured income for Argentina in 1980. Figure 2 corresponds to Mexico in 1989. The average fraction of income reported for Argentina and Mexico, respectively, was .539 and .904. Not surprisingly, a higher average fraction of income reported leads to a closer agreement between the pdf's for actual and measured income.

4. Conclusion

It seems clear that measurement error poses serious problems for economists interested in understanding the trends in inequality over time within a country, or in making cross-country comparisons of income inequality. The problem has huge implications for the formulation of national and international strategies to deal with inequality. If we cannot be confident that the trends in inequality we observe are even in the same direction as the trends in the underlying actual distribution of income, we have little hope of devising programs to influence inequality. In fact, policymakers in countries that have large and variable underreporting of income may be promulgating policies which exacerbate the very problems they believe they are solving. It is hoped that the methodology of this study will provide a means of solving the problem of misleading income inequality measurement.

References

- Arnold, B.C.. 1980. Misreporting Increases Apparent Inequality, University of California at Riverside: mimeographed manuscript.
- Israelsen, L.D., J.B. McDonald, and W.K. Newey. 1984. "The Impact of Measurement Error on the Distribution of Income," **Advances in Econometrics**, 3: 169-189.
- McDonald, J.B. 1984. "Some Generalized Functions for the Size Distribution of Income," **Econometrica**, 52: 647-663.
- McDonald, J.B. and Y.J. Xu. 1995. "A Generalization of the Beta Distribution with Applications," **Journal of Econometrics**, 66: 133-152. Errata 69 (1995), 427-8.
- Psacharopoulos, G., S. Morley, A. Fiszbein, H. Lee, and B. Wood. 1993. "Poverty and Income Distribution in Latin America: The Story of the 1980s," World Bank: Latin America and the Caribbean Technical Department, Report 27.
- Rainville, E.D. 1960. **Special Functions**, New York: MacMillan
- Taillie, C. 1981. "Lorenz Ordering within the Generalized Gamma Family of Income Distributions. " In C. Taillie, G.P. Patil, and B.A. Balderssari (eds), **Statistical Distributions in Scientific Work**, vol. 6: 181-192, London: Reidel.

Table 1

**Measures of Urban* Income Inequality for Selected Latin American Countries, 1980-1989:
Measured Income (Y) and Actual Income (Y*)**

Country	Year	Y Gini Ratio	Y* Gini Ratio	IEF	$\bar{\lambda}$	$\frac{GY^*}{GY}$
Argentina	1980	.425	.288	1.8565	.5386	.678
	1989	.522	.297	2.4508	.4080	.569
Columbia	1980	.570	.322	2.0676	.4837	.565
	1989	.595	.279	1.9355	.5167	.469
Costa Rica	1981	.501	.379	2.4046	.4159	.756
	1989	.485	.319	3.7167	.2691	.658
Guatemala	1987	.619	.328	1.4607	.6846	.530
	1989	.581	.414	1.1231	.8904	.713
Mexico	1984	.536	.272	1.3947	.7170	.507
	1989	.524	.279	1.1068	.9035	.532
Uruguay	1981	.443	.283	1.3835	.7228	.639
	1989	.406	.271	1.3660	.7321	.667
Venezuela	1981	.454	.290	1.6085	.6217	.638
	1989	.459	.287	2.0020	.4995	.625

Y Gini Ratio (GY) is calculated from (2.10).

Y* Gini Ratio (GY*) is calculated from (2.5).

IEY = mean value of the ratio of actual income to measured income ($1/\bar{\lambda}$).

*The data for Mexico 1989 correspond to national income data rather than urban income.

Table 2**Percent Change in Measured and Actual Gini Ratios, 1980-1989**

Country	Time Period	Percent Change in Gy	Percent Change in Gy*
Argentina	1980-1989	22.8	3.13
Columbia	1980-1989	4.4	-13.4
Costa Rica	1981-1989	-3.2	-15.8
Guatemala	1987-1989	-6.1	26.2
Mexico	1984-1989	-2.2	2.6
Uruguay	1981-1989	-8.4	-4.2
Venezuela	1981-1989	1.1	-1.0

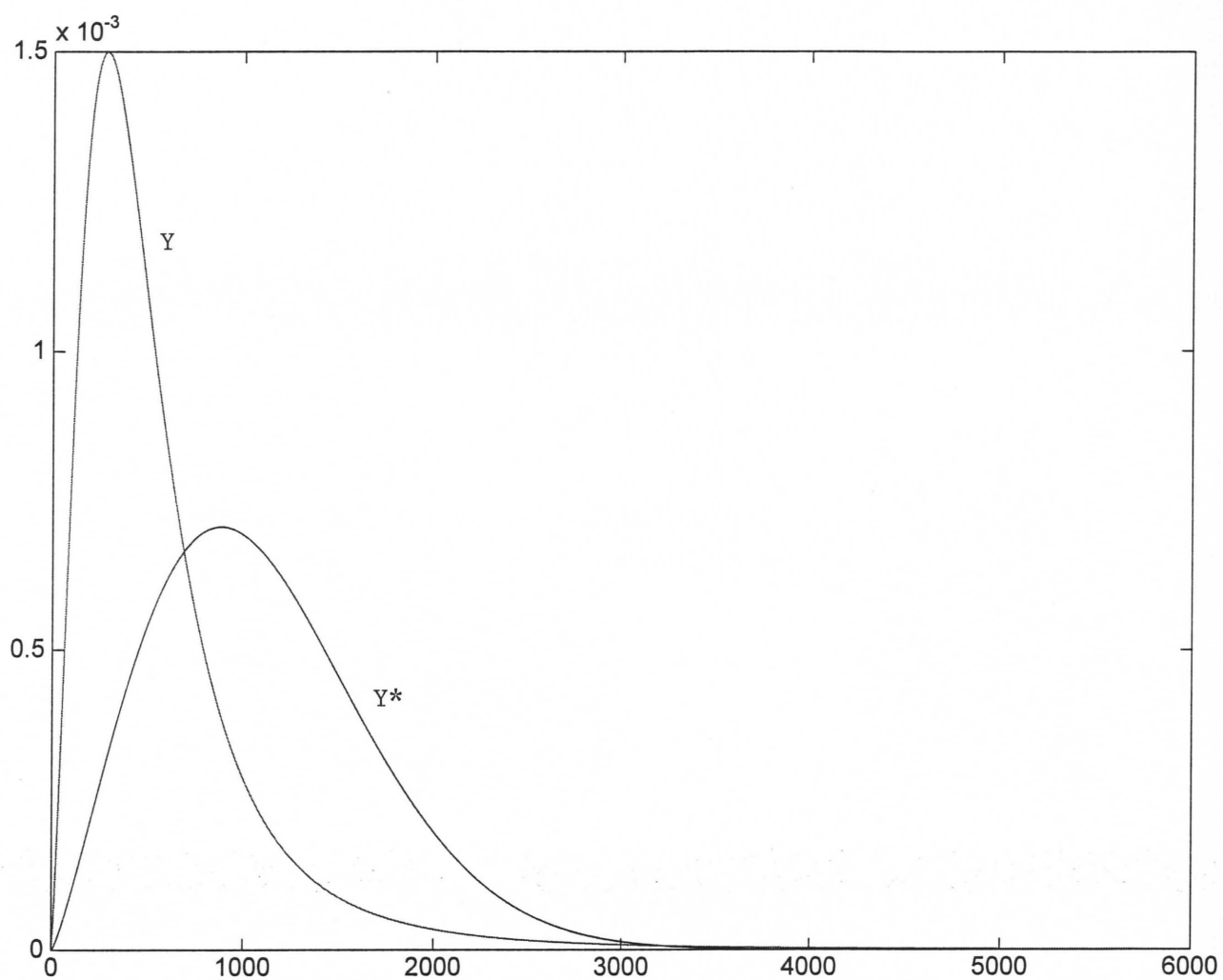


Figure 1

Distributions of Measured (Y) and Actual (Y^*) Income. Argentina, 1980.

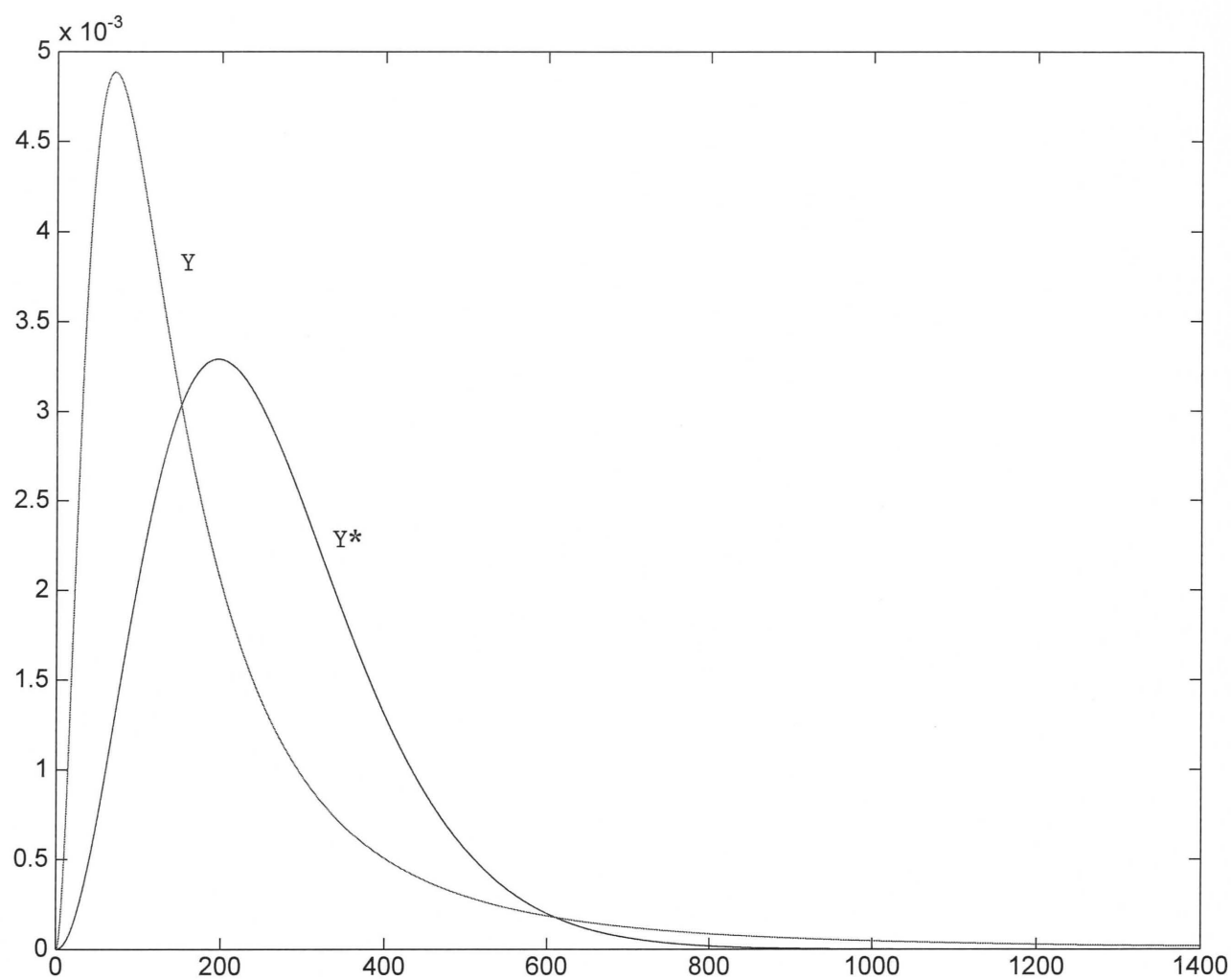


Figure 2

Distributions of Measured (Y) and Actual (Y*) Income. Mexico, 1989.

Measurement Error and the Distribution of Income

L. Dwight Israelsen
Utah State University

James B. McDonald
Brigham Young University

May 2000

The authors appreciate Nora Lustig for her encouragement and for providing the data used in this paper. The comments of Mark Showalter are also greatly appreciated.

Abstract

Measurement Error and the Distribution of Income

L. Dwight Israelsen

James B. McDonald

A model for measurement error is developed, based on the assumption that measurement error is random, multiplicative, and independent of the level of actual income. Thus, measured income is defined as the product of actual income and measurement error. Flexible parametric forms are utilized to model the distributions of actual income (generalized gamma) and measurement error (inverse generalized gamma). The probability density of measured income is then derived as a generalized beta of the second kind (GB2). Estimation of the parameters of the GB2 (measured income), then allows an estimate to be made of the pdf of actual income, from which corresponding estimated means, variances, Gini coefficients and Lorenz curves are obtained. An identification problem in the parameters of actual and measured income is solved with additional information as to the average fraction of actual income reported. The implied characteristics of measurement error are also obtained. The procedure is applied to income data from several Latin American economies, and estimates of actual income distribution characteristics are derived from measured income. It is found, in some cases, that when measured income inequality moved in one direction over time, actual income inequality moved in the opposite direction. This finding has important implications for the evaluation of policies designed to affect relative equality in the distribution of income.

Measurement Error and the Distribution of Income

L. Dwight Israelsen
James B. McDonald

1. Introduction

During the past twenty years, there has been a renewed interest in the United States and around the world in the measurement of income inequality. In the United States, this development has been due, in part, to the uncertainty surrounding the trend in income inequality in the economy since the 1970s. Part of the uncertainty is a result of the recognition by economists and policymakers that the data from which income distributions are estimated is subject to measurement error. The measurement error has generally been assumed to be associated with underreporting; however, during the past decade concern has also been expressed about the problem of measurement error/evasion due to overreporting of income by low income families wanting to claim the earned income tax credit.

The purpose of the research reported here is to develop a tractable method to estimate the actual distribution of income from the distribution of measured income in the presence of measurement error with certain characteristics. Estimation of the parameters of the distribution of measured income would then allow us to recover an estimate of the probability density function of actual income. From the pdf of actual income, we could then obtain corresponding estimated means, variances, Gini coefficients and Lorenz curves.

This paper expands and improves on our earlier paper (Israelsen, McDonald & Newey, 1984) in which we derived several pdf's for measured income and demonstrated the theoretical possibilities of the methodology. Unfortunately, the distribution functions we derived there were

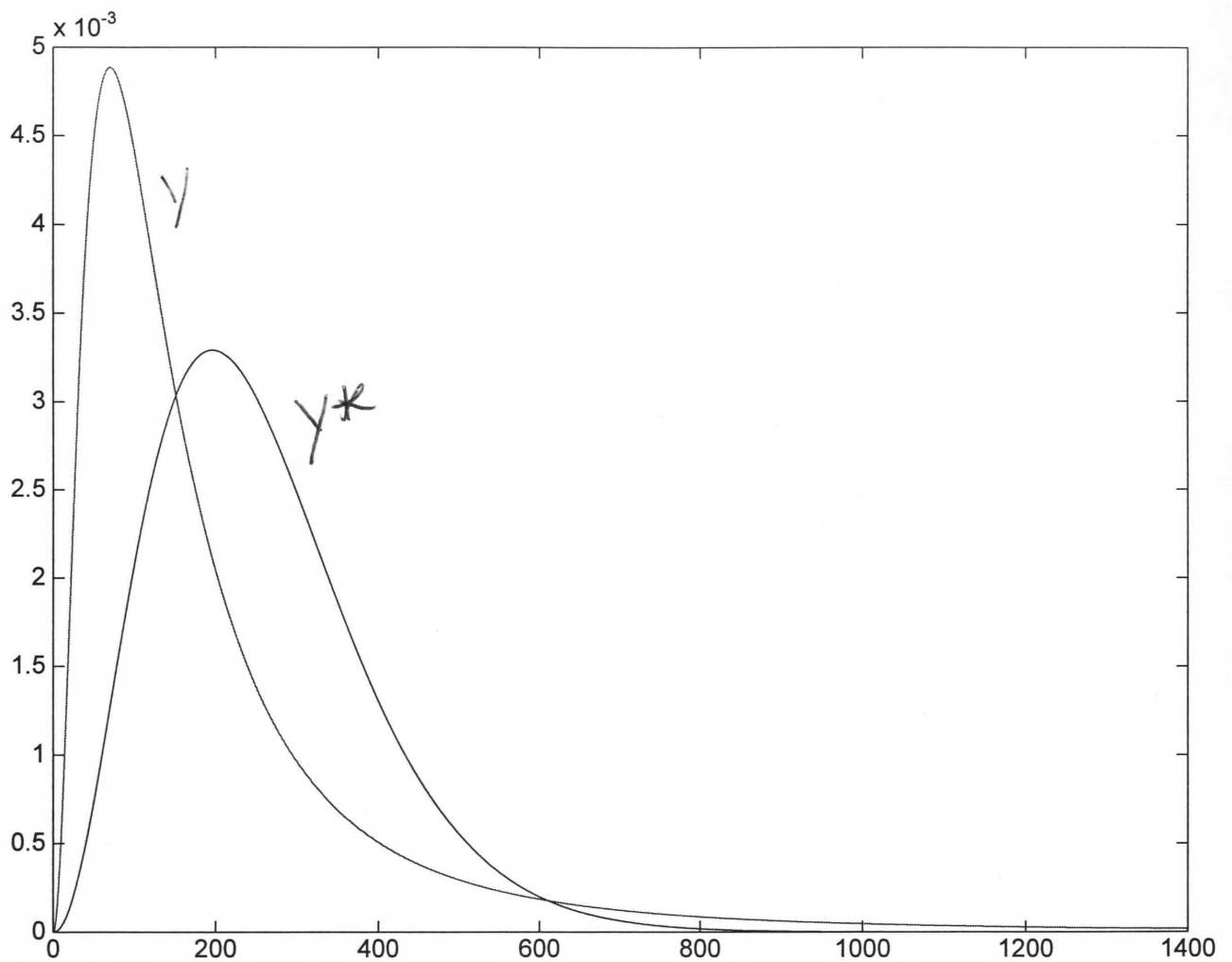


Figure 2. Distributions of Measured (Y) and Actual (Y^*) Income. Mexico, 1989.

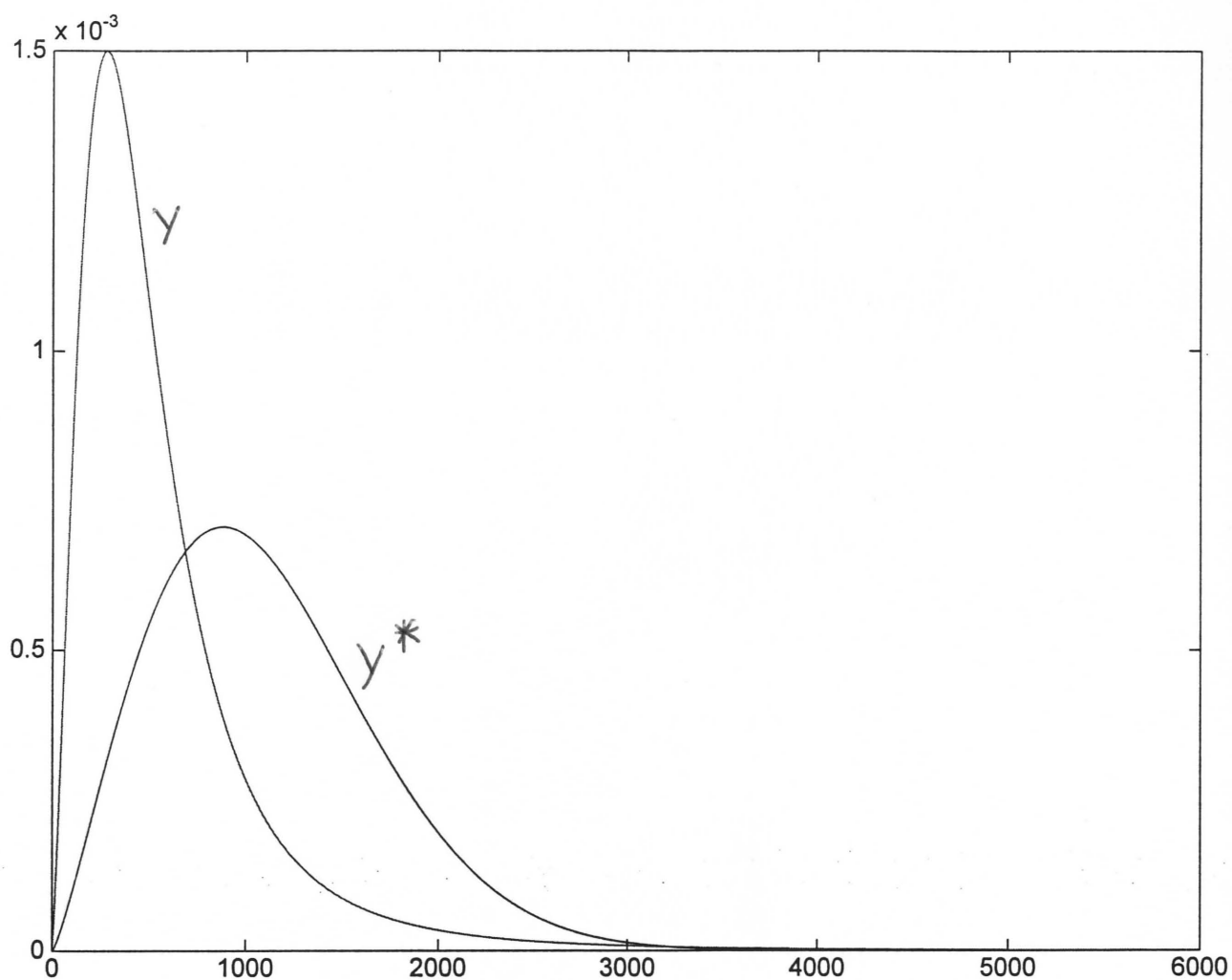


Figure 1. Distributions of Measured (Y) and Actual (Y^*) Income. Argentina, 1980